ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

J. Math. Anal. Appl. 341 (2008) 467–477

Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS

www.elsevier.com/locate/jmaa

Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds

E.A. Papa Quiroz, E.M. Quispe, P. Roberto Oliveira *

Department of Systems Engineering and Computer Science, PESC-COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Received 25 May 2006

Available online 11 October 2007

Submitted by Steven G. Krantz

Abstract

This paper extends the full convergence of the steepest descent method with a generalized Armijo search and a proximal regularization to solve minimization problems with quasiconvex objective functions on complete Riemannian manifolds. Previous convergence results are obtained as particular cases and some examples in non-Euclidian spaces are given. In particular, our approach can be used to solve constrained minimization problems with nonconvex objective functions in Euclidian spaces if the set of constraints is a Riemannian manifold and the objective function is quasiconvex in this manifold.

Keywords: Steepest descent method; Riemannian manifolds; Quasiconvex functions

1. Introduction

The steepest descent method is one of the oldest and most widely known methods in the literature for solving smooth optimization problems. However, classical results for an arbitrary objective function, are not very strong because the full convergence and the existence of cluster points are not assured. We only may guarantee that any cluster point, if it exists, is a critical point of the problem. The situation is very different when the objective function is convex, because under the only assumption of existence of the optimal set, the steepest descent method with Armijo search and a proximal regularization converges to an optimal point, see Burachik et al. [4] and Iusem and Svaiter [15]. The first result was recently generalized to finite dimensional vector spaces by Graña Drummond and Svaiter [12].

On the other hand, generalization of optimization methods from Euclidian spaces to Riemannian manifolds have some important advantages. For example, constrained optimization problems can be seen as unconstrained ones from the Riemannian geometry point of view, another advantage is that optimization problems with nonconvex objective functions become convex through the introduction of a appropriate Riemannian metric, see for example [24].

The steepest descent method to solve optimization problems for arbitrary objective functions on finite dimensional Riemannian manifolds has been studied by Udriste [24], Smith [22] and Rapcsák [21], where they have obtained the

* Corresponding author.
E-mail addresses: erik@cos.ufrj.br (E.A. Papa Quiroz), mariss@cos.ufrj.br (E.M. Quispe), poliveir@cos.ufrj.br (P.R. Oliveira).

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2007.10.010

same weakly convergence results of the classical method. For the convex case on those manifolds but with nonnegative sectional curvature, the full convergence using Armijo search, fixed steps and a proximal regularization, has been generalized by Cruz Neto et al. [5,6].

In this paper, we generalize the full convergence of the steepest descent method on such manifolds for the quasiconvex case following the ideas of Kiwiel and Murty [16] but using the theory of quasi-Fejér convergence. An important characteristic of our approach is that it recovers all the previous results of the convex case. Indeed, we prove with the only assumption that the optimal set is nonempty the full convergence of that method to a critical point of the problem.

Our motivation to study this subject comes from two fields. One of them, is the broad range of applications of quasiconvex optimization in diverse areas of sciences and engineering. For example, in economic theory [23], location theory [13], control theory [1] and dynamical systems [11]. In this context, if the constraints of a minimization problem constitute a Riemannian manifold, and the objective function is quasiconvex, the problem becomes an unconstrained one, and therefore it is not necessary to make projections in each iteration of the steepest descent method. The other motivation, is to solve more general optimization problems with nonconvex objective functions. Now, the Riemannian techniques can transform those problems in quasiconvex ones, under appropriate metrics on the manifolds. We believe that this kind of application could permit more examples, if we compare with the convex case. For the interested reader in the literature on convex and optimization problems, we refer to [2,3]. On the other hand, we point out that an important class of nonconvex problems is given by nonconvex quadratic problems (with SDP relaxations as a possible issue), as can be seen in [14, Chapter 13].

The paper is divided as follows. In Section 2 we give some results of metric spaces and Riemannian geometry that we will use along the paper. In Section 3, we analyze the steepest descent method for the quasiconvex case on Riemannian manifolds with nonnegative sectional curvature. We prove the full convergence of this method to a critical point of the problem using a generalized Armijo search and a proximal regularization. In Section 4 we give some examples of steepest descent methods to solve optimization problems with nonconvex objective functions in Euclidian spaces. Finally, in Section 5 we give some numerical experiments to illustrate the main result of our approach.

2. Some basic facts on metric and Riemannian spaces

Definition 2.1. Let (X, d) be a complete metric space. A sequence $\{y^k\}$, $k \geqslant 0$, of X is *quasi-Fejér convergent* to a set $U \subset X$, if for every $u \in U$ there exists a sequence $\{\epsilon_k\} \subseteq \mathbb{R}$ such that $\epsilon_k \geqslant 0$, $\sum_{k=0}^{+\infty} \epsilon_k < +\infty$ and

$$d^2\big(y^{k+1}, u\big) \leqslant d^2\big(y^k, u\big) + \epsilon_k.$$

Theorem 2.1. *In a complete metric space (X, d), if $\{y^k\}$ is quasi-Fejér convergent to a nonempty set $U \subseteq X$, then $\{y^k\}$ is bounded. If, furthermore, a cluster point $\bar{y}$ of $\{y^k\}$ belongs to U, then $\{y^k\}$ converges and* $\lim_{k\to+\infty} y^k = \bar{y}$.

Proof. Analogous to Burachik et al. [4] replacing the Euclidian norm by the distance d. □

Now, we recall some fundamental properties and notation on Riemannian manifolds. Those basic facts can be seen, for example, in do Carmo [8].

Throughout this paper all manifolds are assumed to be smooth and connected. Let M be a finite dimensional manifold, we denote by $T_x M$ the tangent space of M at x and $TM = \bigcup_{x\in M} T_x M$. $T_x M$ is a linear space and has the same dimension of M. Because we restrict ourselves to real manifolds, $T_x M$ is isomorphic to $\mathbb{R}^n$. If M is endowed with a Riemannian metric g, then M is a Riemannian manifold and we denoted it by (M, G) or only by M when no confusion can arise, where G denotes the matrix representation of the metric g. The inner product of two vectors $u, v \in T_x S$ is written $\langle u, v\rangle_x := g_x(u, v)$, where g_x is the metric at the point x. The norm of a vector $v \in T_x S$ is defined by $\|v\|_x := \langle v, v\rangle_x^{1/2}$. The metric can be used to define the length of a piecewise smooth curve $\alpha : [t_0, t_1] \to S$ joining $\alpha(t_0) = p'$ to $\alpha(t_1) = p$ through $L(\alpha) = \int_{t_0}^{t_1} \|\alpha'(t)\|\, dt$. Minimizing this length functional over the set of all curves we obtain a Riemannian distance $d(p', p)$ which induces the original topology on M.

Given two vector fields V and W in M (a vector field V is an application of M in TM), the covariant derivative of W in the direction V is denoted by $\nabla_V W$. In this paper ∇ is the Levi-Civita connection associated to (M, G). This connection defines an unique covariant derivative D/dt, where for each vector field V, along a smooth curve

$\alpha : [t_0, t_1] \to M$, another vector field is obtained, denoted by DV/dt. The parallel transport along α from $\alpha(t_0)$ to $\alpha(t_1)$, denoted by P_{α,t_0,t_1}, is an application $P_{\alpha,t_0,t_1} : T_{\alpha(t_0)}M \to T_{\alpha(t_1)}M$ defined by $P_{\alpha,t_0,t_1}(v) = V(t_1)$ where V is the unique vector field along α such that $DV/dt = 0$ and $V(t_0) = v$. Since that ∇ is a Riemannian connection, P_{α,t_0,t_1} is a linear isometry, furthermore $P^{-1}_{\alpha,t_0,t_1} = P_{\alpha,t_1,t_0}$ and $P_{\alpha,t_0,t_1} = P_{\alpha,t,t_1}P_{\alpha,t_0,t}$, for all $t \in [t_0, t_1]$. A curve $\gamma : I \to M$ is called a geodesic if $D\gamma'/dt = 0$. A geodesic curve γ, starting from the point x with direction $v \in T_xS$ ($\gamma(0) = x$, $\gamma'(0) = v$) is given by

$$\frac{d^2\gamma_k}{dt^2} + \sum_{i,j=1}^{n} \Gamma^k_{ij}\frac{d\gamma_i}{dt}\frac{d\gamma_j}{dt} = 0, \quad k = 1, \dots, n,$$

where Γ^k_{ij} are the Christoffel's symbols expressed by

$$\Gamma^m_{ij} = \frac{1}{2}\sum_{k=1}^{n}\left\{\frac{\partial}{\partial x_i}g_{jk} + \frac{\partial}{\partial x_j}g_{ki} - \frac{\partial}{\partial x_k}g_{ij}\right\}g^{km},$$

(g^{ij}) denotes the inverse matrix of the metric $g = (g_{ij})$, and x_i is the coordinates of x. A Riemannian manifold is complete if its geodesics are defined for any value of $t \in \mathbb{R}$. Let $x \in M$, the exponential map $\exp_x : T_xM \to M$ is defined as $\exp_x(v) = \gamma(1)$. If M is complete, then $\exp_x$ is defined for all $v \in T_xM$. Besides, there is a minimal geodesic (its length is equal to the distance between the extremes).

Given the vector fields X, Y, Z on M, we denote by R the curvature tensor defined by $R(X,Y)Z = \nabla_Y\nabla_XZ - \nabla_X\nabla_YZ + \nabla_{[X,Y]}Z$, where $[X,Y] := XY - YX$ is the Lie bracket. Now, the sectional curvature with respect to X and Y is defined by

$$K(X,Y) = \frac{\langle R(X,Y)Y, X\rangle}{\|X\|^2\|Y\|^2 - \langle X,Y\rangle^2}.$$

If $K(X,Y) \geqslant 0$, then M is a Riemannian manifold with nonnegative sectional curvature. The gradient of a differentiable function $f : M \to \mathbb{R}$, $\operatorname{grad} f$, is a vector field on M defined through $df(X) = \langle \operatorname{grad} f, X\rangle = X(f)$, where X is also a vector field on M.

A geodesic hinge in M is a pair of normalized geodesics segments γ_1 and γ_2 such that $\gamma_1(0) = \gamma_2(0)$ and at least one of them, say γ_1, is minimal.

Theorem 2.2. *In a complete finite dimensional Riemannian manifold M with nonnegative sectional curvature we have*

$$l_3^2 \leqslant l_1^2 + l_2^2 - 2l_1l_2\cos\alpha,$$

where l_i denote the length of γ_i ($i = 1, 2$), $l_3 = d(\gamma_1(l_1), \gamma_2(l_2))$ and $\alpha = \angle(\gamma_1'(0), \gamma_2'(0))$.

Proof. See [5, Theorem 2.1]. □

Definition 2.2. Let M be a Riemannian manifold and $f : M \to \mathbb{R}$ be a real function. f is called quasiconvex on M if for all $x, y \in M$, $t \in [0, 1]$, it holds that

$$f\big(\gamma(t)\big) \leqslant \max\big\{f(x), f(y)\big\},$$

for all geodesic $\gamma : [0, 1] \to \mathbb{R}$, such that $\gamma(0) = x$ and $\gamma(1) = y$.

Theorem 2.3. *Let $f : M \to \mathbb{R}$ be a differentiable quasiconvex function on a complete Riemannian manifold M and let $x, y \in M$. If $f(x) \leqslant f(y)$ then*

$$\big\langle \operatorname{grad} f(y), \gamma'(0)\big\rangle \leqslant 0,$$

where $\operatorname{grad} f$ is the gradient of f and γ is a geodesic curve such that $\gamma(0) = y$ and $\gamma(1) = x$.

Proof. See Németh [17, Proposition 3.1]. □

Definition 2.3. A differentiable function $f : M \to \mathbb{R}$ is pseudoconvex if, for every pair of distinct points x, y and every geodesic curve joining x to y ($\gamma(0) = x$ and $\gamma(1) = y$) we have

$$\langle \mathtt{grad}\, f(x), \gamma'(0)\rangle \geqslant 0, \quad \text{then } f(y) \geqslant f(x).$$

Theorem 2.4. *Let $f : M \to \mathbb{R}$ be a differentiable pseudoconvex function. Then, x^* is a global minimum of f if, and only if,* $\mathtt{grad}\, f(x^*) = 0$.

Proof. It is immediate. □

3. The steepest descent method

We are interested in solving the optimization problem

$$(p) \min_{x\in M} f(x),$$

where M is a complete finite dimensional Riemannian manifold and $f : M \to \mathbb{R}$ is a continuously differentiable quasiconvex function. The steepest descent method generates a sequence $\{x^k\}$ given by

$$x^0 \in M, \tag{1}$$

$$x^{k+1} = \exp_{x^k}\left(-t_k\, \mathtt{grad}\, f\left(x^k\right)\right), \tag{2}$$

where exp is the exponential map and t_k is some positive stepsize. We first assume the following:

Assumption A1. The global optimal set of the problem (p), denoted by X^*, is nonempty. We denote the optimal value of (p) by f^*. Now, we define the following nonempty set:

$$U := \left\{x \in M\colon\ f(x) \leqslant \inf_k f\left(x^k\right)\right\}.$$

The following lemma is the key of our paper because we will used this fact to prove that the sequence, defined by the steepest descent method, is quasi-Fejér convergent to U.

Lemma 3.1. *Let $f : M \to \mathbb{R}$ be a continuously differentiable quasiconvex function on a complete finite dimensional Riemannian manifold with nonnegative sectional curvature, then*

$$d^2\left(x^{k+1}, x\right) \leqslant d^2\left(x^k, x\right) + t_k^2 \left\|\mathtt{grad}\, f\left(x^k\right)\right\|^2$$

for all $x \in U$ and all $t_k > 0$.

Proof. Let $x \in U$ arbitrary. Suppose $\gamma_1 : [0, 1] \to M$ a minimal geodesic segment linking x^k to x, and $\gamma_2 : [0, 1] \to M$ the geodesic segment linking x^k to x^{k+1} with $\gamma_2'(0) = -t_k\, \mathtt{grad}\, f(x^k)$. From Theorem 2.2 we have that

$$d^2\left(x^{k+1}, x\right) \leqslant d^2\left(x^k, x\right) + t_k^2 \left\|\mathtt{grad}\, f\left(x^k\right)\right\|^2 + 2t_k d\left(x^k, x\right)\left\langle \mathtt{grad}\, f\left(x^k\right), \gamma_1'(0)\right\rangle.$$

As f is quasiconvex and $f(x) \leqslant f(x^k)$, from Theorem 2.3 we obtain that $\langle \mathtt{grad}\, f(x^k), \gamma_1'(0)\rangle \leqslant 0$. Using this fact in the previous inequality, we conclude the proof. □

From now on M will be a *complete finite dimensional Riemannian manifold with nonnegative sectional curvature*.

3.1. The steepest descent method with a generalized Armijo search

The steepest descent method with Armijo's stepsize generates $\{x^k\}$ given by (1)–(2) where

$$t_k = \arg\max\left\{t\colon\ f\left(\exp_{x^k}\left(-t\, \mathtt{grad}\, f\left(x^k\right)\right)\right) \leqslant f\left(x^k\right) - \alpha t \left\|\mathtt{grad}\, f\left(x^k\right)\right\|^2,\ t = 2^{-i},\ i = 0, 1, \ldots\right\} \tag{3}$$

with $\alpha \in (0, 1)$.

In this subsection we prove the full convergence of this method for the quasiconvex case. Our results are a generalization of Kiwiel and Murty [16] to the Riemannian framework and extend previous convergence results obtained, for the convex case, by Burachik et al. [4] and Cruz Neto et al. [5]. As in [16], consider the following assumption:

Assumption A2. Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a function such that

A2.1 There exist $\alpha \in (0,1)$, $\tau_\alpha > 0$, such that $\forall t \in (0, \tau_\alpha]$: $\phi(t) \leqslant \alpha t$.
A2.2 There exist $\beta > 0$, $\tau_\beta \in (0, +\infty]$, such that $\forall t \in (0, \tau_\beta) \cap \mathbb{R}$: $\phi(t) \geqslant \beta t^2$.
A2.3 For all k, $f(x^{k+1}) \leqslant f(x^k) - \phi(t_k)\|\operatorname{grad} f(x^k)\|^2$ and $0 < t_k \leqslant \tau_\beta$ in (2).
A2.4 There exist $\gamma > 1$, $\tau_\gamma > 0$, such that $\forall k$: $t_k \geqslant \tau_\gamma$ or

$$\left[\text{there exists } \bar{t}_k \in [t_k, \gamma t_k]\text{: } f\left(\exp_{x^k}\left(-\bar{t}_k \operatorname{grad} f\left(x^k\right)\right)\right) \geqslant f\left(x^k\right) - \phi(\bar{t}_k)\left\|\operatorname{grad} f\left(x^k\right)\right\|^2\right].$$

Remark 3.1. We point out that Assumption A2 is satisfied by the Armijo rule (3) to $\phi(t) = \alpha t$, $\beta = \alpha$, $\gamma = 2$ and $\tau_\alpha = \tau_\beta = \tau_\gamma = 1$.

Remark 3.2. Assumption A2 also is satisfied by the steepest descent method with fixed step introduced in [4] and generalized to Riemannian manifolds by [5]. Of fact, in [4] and [5] the rule to obtain t_k is the following:

Given δ_1 and δ_2 such that $\delta_1 \Gamma + \delta_2 < 1$, where Γ is the Lipschitz constant associated to $\operatorname{grad} f$, choose

$$t_k \in \left(\delta_1, \frac{2}{\Gamma}(1 - \delta_2)\right).$$

Now, defining $\phi(t) = \beta t^2$, with $\beta = \frac{\Gamma \delta_2}{2(1-\delta_2)}$, $\tau_\gamma = \delta_1$, $\tau_\beta = (2/\Gamma)(1 - \delta_2)$, $\alpha \in (0,1)$ arbitrary and $\tau_\alpha = \alpha/\beta$, we assure Assumption A2.

Proposition 3.1. *Let $f : M \to \mathbb{R}$ be a continuously differentiable quasiconvex function. Suppose that Assumptions* A1 *and* A2 *are satisfied. Then, the sequence $\{x^k\}$ generates by the steepest descent method with generalized Armijo search is quasi-Fejér convergent to U.*

Proof. From Assumptions A2.2 and A2.3 we have

$$\beta t_k^2 \left\|\operatorname{grad} f\left(x^k\right)\right\|^2 \leqslant f\left(x^k\right) - f\left(x^{k+1}\right). \tag{4}$$

This implies that

$$\sum_{k=0}^{+\infty} t_k^2 \left\|\operatorname{grad} f\left(x^k\right)\right\|^2 \leqslant \frac{f(x^0) - f^*}{\beta} < +\infty.$$

From Lemma 3.1 and Definition 2.1 we obtain the result. □

Theorem 3.1. *Let $f : M \to \mathbb{R}$ be a continuously differentiable quasiconvex function. Suppose that Assumptions* A1 *and* A2 *are satisfied. Then the sequence $\{x^k\}$ generates by the steepest descent method with generalized Armijo search converges. Moreover, it converges to a stationary point* (*a point $\bar{x}$ such that* $\operatorname{grad} f(\bar{x}) = 0$).

Proof. From previous proposition, $\{x^k\}$ is Fejér convergent to U, thus $\{x^k\}$ is bounded (see Theorem 2.1). Then, there exist $\bar{x}$ and a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to $\bar{x}$. From continuity of f we obtain

$$\lim_{j \to +\infty} f\left(x^{k_j}\right) = f(\bar{x}).$$

As $\{f(x^k)\}$ is a nonincreasing sequence, see (4), with a subsequence converging to $f(\bar{x})$, the overall sequence converges to $f(\bar{x})$ and therefore

$$f(\bar{x}) \leqslant f\left(x^k\right), \quad \forall k \in \mathbb{N}.$$

This implies that $\bar{x} \in U$. Now, from Theorem 2.1 we can conclude that $\{x^k\}$ converges to $\bar{x}$. Finally, we will prove that $\operatorname{grad} f(\bar{x}) = 0$. By contradiction, we suppose that $\operatorname{grad} f(\bar{x}) \neq 0$. Clearly, we have $\operatorname{grad} f(x^k) \to \operatorname{grad} f(\bar{x}) \neq 0$ and $f(x^k) \to f(\bar{x})$. Now, from (4) it holds that

$$\lim_{k \to +\infty} t_k = 0. \tag{5}$$

On the other hand, using A2.4 and A2.1 we have, for k large enough,

$$f\big(\exp_{x^k}\big(-\bar{t}_k \operatorname{grad} f\big(x^k\big)\big)\big) - f\big(x^k\big) \geqslant -\alpha \bar{t}_k \big\|\operatorname{grad} f\big(x^k\big)\big\|^2. \tag{6}$$

Besides, from the mean value theorem, for such k, there exists $t_k^* \in [0, \bar{t}_k]$ such that

$$-\big\langle \operatorname{grad} f\big(\exp_{x^k}\big(-t_k^* \operatorname{grad} f\big(x^k\big)\big)\big), P_{\gamma_k,0,t_k^*} \operatorname{grad} f\big(x^k\big)\big\rangle \geqslant -\alpha \big\|\operatorname{grad} f\big(x^k\big)\big\|^2,$$

where $P_{\gamma_k,0,t_k^*}$ is the parallel transport along the geodesic γ_k such that $\gamma_k(0) = x^k$ and $\gamma_k'(0) = -\operatorname{grad} f(x^k)$. Now, (5) and A2.4 imply $\lim_{k\to+\infty} t_k^* = 0$. Letting $k \to \infty$ in the previous inequality and taking in account the continuity of $\operatorname{grad} f$, exp and parallel transport, we have that $1 \leqslant \alpha$, which is a contradiction with A2.1. Therefore $\operatorname{grad} f(\bar{x}) = 0$. □

As a consequence of the previous theorem and Theorem 2.4 we have the following result:

Corollary 3.1. *Let $f : M \to \mathbb{R}$ be a pseudoconvex function. Then, under Assumptions* A1 *and* A2*, the sequence $\{x^k\}$ converges to a global minimum of (p).*

3.2. The steepest descent method with a proximal regularization

Let $\{\lambda_k\}$ be a real sequence such that $\lambda' \leqslant \lambda_k \leqslant \lambda''$, where $0 < \lambda' \leqslant \lambda''$. The regularized steepest decent method generates a sequence $\{x^k\}$ defined by (1)–(2), where

$$t_k = \arg\min\big\{t \geqslant 0\colon\ f\big(\exp_{x^k}\big(-t \operatorname{grad} f\big(x^k\big)\big)\big) + t^2 \lambda_k \big\|\operatorname{grad} f\big(x^k\big)\big\|^2\big\}. \tag{7}$$

This method has been introduced by Iusem and Svaiter [15] to solve convex optimization problems on Euclidian spaces and then generalized to the Riemannian manifolds by Cruz Neto et al. [6]. In this subsection we extend the full convergence results of those works to the quasiconvex case.

Proposition 3.2. *Let $f : M \to \mathbb{R}$ be a continuously differentiable quasiconvex function. Suppose that Assumption* A1 *is satisfied. Then the sequence $\{x^k\}$, generated by* (1), (2) *and* (7)*, is quasi-Fejér convergent to the set U.*

Proof. From (2) and (7):

$$f\big(x^{k+1}\big) + t_k^2 \lambda_k \big\|\operatorname{grad} f\big(x^k\big)\big\|^2 \leqslant f\big(x^k\big). \tag{8}$$

Hence, it is easy to verify that

$$\sum_{k=0}^{+\infty} t_k^2 \big\|\operatorname{grad} f\big(x^k\big)\big\|^2 \leqslant (1/\lambda')\big(f\big(x^0\big) - f^*\big) < +\infty.$$

From Lemma 3.1 and Definition 2.1, we conclude the proof. □

Theorem 3.2. *Let $f : M \to \mathbb{R}$ be a continuously differentiable quasiconvex function. Suppose that Assumption* A1 *is satisfied. Then the sequence $\{x^k\}$, generates by* (1), (2) *and* (7)*, converges to a stationary point.*

Proof. From (8) we have that $\{f(x^k)\}$ is a nonincreasing sequence. Using the same arguments of the proof of Theorem 3.1, we can show that $\{x^k\}$ converges to a point $x^* \in U$. Finally, we have $\operatorname{grad} f(x^*) = 0$, as an application of Theorem 4.1, iiii, in [6], where it is shown that property for an arbitrary function. □

Similarly to Corollary 3.1 we have

Corollary 3.2. *Let $f : M \to \mathbb{R}$ be a pseudoconvex function on a complete finite dimensional Riemannian manifold M with nonnegative sectional curvature. Then, under Assumption* A1*, the sequence $\{x^k\}$ converges to a global minimum point of (p).*

4. Some examples

The following algorithms solve, in particular, minimization problems with nonconvex objective functions in Euclidian spaces.

When necessary, we let X denoting the diagonal matrix $X = \operatorname{diag}(x_1, \dots, x_n)$.

4.1. A steepest descent algorithm for $\mathbb{R}^n$

Consider the problem

$$\min\{f(x):\ x \in \mathbb{R}^n\}.$$

Take $\mathbb{R}^n$ and consider the metric

$$G(x) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 1+4x_{n-1}^2 & -2x_{n-1} \\ 0 & \cdots & 0 & 0 & 0 & -2x_{n-1} & 1 \end{bmatrix}.$$

Thus $(\mathbb{R}^n, G(x))$ is a connected and complete finite dimensional Riemannian manifold with null sectional curvature, see [7]. The gradient of f is given by $\operatorname{grad} f(x) = G^{-1}(x)\nabla f(x)$ and the steepest descent iteration is

$$x_i^{k+1} = x_i^k - t_k \frac{\partial f(x^k)}{\partial x_i}, \quad \forall i = 1, \dots, n-2,$$

$$x_{n-1}^{k+1} = x_{n-1}^k - t_k \left(\frac{\partial f(x^k)}{\partial x_{n-1}} + 2x_{n-1}^k \frac{\partial f(x^k)}{\partial x_n} \right),$$

$$x_n^{k+1} = x_n^k - t_k \left(2x_{n-1}^k \frac{\partial f(x^k)}{\partial x_{n-1}} + \left(1 + 4\left(x_{n-1}^k\right)^2\right) \frac{\partial f(x^k)}{\partial x_n} \right).$$

4.2. A steepest descent algorithm for $\mathbb{R}^n_{++}$

Consider the problem

$$\min\{f(x):\ x \geqslant 0\}.$$

Take the positive octant $\mathbb{R}^n_{++}$ and consider the Riemannian manifold $(\mathbb{R}^n_{++}, X^{-2})$ (X^{-2} is the Hessian of the $-\log$ barrier). This space is a connected and complete finite dimensional Riemannian manifold with null sectional curvature. The gradient of f is given by $\operatorname{grad} f(x) = X^2 \nabla f(x)$ (the opposite of the affine scaling direction) and the steepest descent iteration is

$$x_i^{k+1} = x_i^k \exp\left(-x_i^k \frac{\partial f(x^k)}{\partial x_i} t_k \right), \quad i = 1, 2, \dots, n.$$

4.3. Steepest descent algorithms for the hypercube

Let the problem

$$\min\{f(x):\ 0 \leqslant x \leqslant e\}.$$

Take $(0, 1)^n$ as a smooth manifold. We will introduce three connected and complete finite dimensional Riemannian manifolds with null sectional curvatures. In these applications, we use capital letters, for example, $\mathrm{CSC}(v) = \operatorname{diag}(\csc(v_1), \dots, \csc(v_n))$, to represent the diagonal matrix composed by the corresponding trigonometric function.

(a) $((0,1)^n, X^{-2}(I-X)^{-2})$. Then $\mathtt{grad}\, f(x) = X^2(I-X)^2\nabla f(x)$. The steepest descent iteration is

$$x_i^{k+1} = \frac{1}{2}\left\{1+\tanh\left(-\frac{1}{2}x_i^k\left(1-x_i^k\right)\frac{\partial f(x^k)}{\partial x_i}t_k + \frac{1}{2}\ln\frac{x_i^k}{1-x_i^k}\right)\right\}, \quad i=1,2,\dots,n.$$

(b) $((0,1)^n, \mathrm{CSC}^4(\pi x))$. Then $\mathtt{grad}\, f(x) = \mathrm{SIN}^4(\pi x)\nabla f(x)$. The steepest descent iteration is

$$x_i^{k+1} = \frac{1}{\pi}\operatorname{arccot}\left(\pi\sin^2\left(\pi x_i^k\right)\frac{\partial f(x^k)}{\partial x_i}t_k + \cot\left(\pi x_i^k\right)\right), \quad i=1,2,\dots,n.$$

(c) $((0,1)^n, \mathrm{CSC}^2(\pi x))$, see Nesterov and Todd [20]. The gradient of f is given by $\mathtt{grad}\, f(x) = \mathrm{SIN}^2(\pi x)\nabla f(x)$ and the steepest descent iteration is

$$x_i^{k+1} = \psi^{-1}\left(-\sin\left(\pi x_i^k\right)\frac{\partial f(x^k)}{\partial x_i}t_k + \psi\left(x_i^k\right)\right), \quad i=1,2,\dots,n,$$

where

$$\psi(z) := (-1/\pi)\ln\left(\csc(\pi z)+\cot(\pi z)\right).$$

Remark 4.1. The metric $X^{-2}(I-X)^{-2}$ is the Hessian of the self-concordant barrier $B(x) = \sum_{i=1}^n (2x_i - 1)[\ln x_i - \ln(1-x_i)]$, exploited in Papa Quiroz and Roberto Oliveira [19].

$\csc^4(\pi x)$, given in Section 4.3(b), is new, to our knowledge. It is the Hessian of a C^∞ strictly convex self-concordant function (see Definition 2.1.1 of Nesterov and Nemirovskii [18]), allowing the introduction of new interior point algorithms for convex optimization problems, as proximal and subgradient. We observe that general convergence results could be applied for those methods, through, respectively, the theory developed in [10] and [9].

5. Numerical experiments

In this section we give some numerical experiments to solve minimization problems with quasiconvex objective functions on the unitary hypercube, that is

$$\min\left\{f(x):\ 0 \leqslant x \leqslant e\right\}, \tag{9}$$

where f is a quasiconvex function, $x = (x_1, x_2, \dots, x_n)$ is the variable of the problem and $e = (1,\dots,1) \in \mathbb{R}^n$.

Take the connected and complete Riemannian manifold $((0,1)^n, X^{-2}(I-X)^{-2})$, then the steepest descent algorithm with Armijo search, see (3), works essentially as follows:

1. Given a point $x^k = (x_1^k, x_2^k, \dots, x_n^k) \in (0,1)^n$, $k \geqslant 0$, compute x^{k+1} given by

$$x_i^{k+1} = \frac{1}{2}\left\{1+\tanh\left(-\frac{1}{2}x_i^k\left(1-x_i^k\right)\frac{\partial f(x^k)}{\partial x_i}t_k + \frac{1}{2}\ln\frac{x_i^k}{1-x_i^k}\right)\right\}, \quad i=1,2,\dots,n,$$

where $t_k = 2^{-i_k}$, i_k is the least positive natural number such that

$$f\left(x^{k+1}\right) \leqslant f\left(x^k\right) - \alpha t_k \left\|d^k\right\|^2,$$

$d^k = -X_k^2(I-X_k)^2\nabla f(x^k)$ is the gradient of f with respect to the metric $X_k^{-2}(I-X_k)^{-2}$, $X_k = \mathrm{diag}(x_1^k, x_2^k, \dots, x_n^k)$, $\nabla f(x^k)$ is the classic gradient of f and $\alpha \in (0,1)$ is given.

2. As a stop criterium we compute the geodesic distance between the points x^k and x^{k+1}, as

$$d\left(x^k, x^{k+1}\right) = \left\{\sum_{i=1}^n\left[\ln\left(\frac{x_i^{k+1}}{1-x_i^{k+1}}\right) - \ln\left(\frac{x_i^k}{1-x_i^k}\right)\right]^2\right\}^{\frac{1}{2}}.$$

3. Stop test: if $\|d(x^k, x^{k+1})\| < \epsilon$, stop. Otherwise, make $x^k \leftarrow x^{k+1}$ and return the step 1.

Table 1

$X0$	Iter.	Call. Armijo	Opt. point	Opt. value	Riem. distance
(0.45, 0.51)	65	65	(0.499999, 0.5)	1.66511	9.27003e−007
(0.4, 0.6)	71	71	(0.499999, 0.500001)	1.66511	9.93398e−007
(0.1, 0.9)	85	85	(0.499999, 0.500001)	1.66511	8.92053e−007
(0.2, 0.3)	79	79	(0.499999, 0.499999)	1.66511	8.79813e−007
(0.7, 0.6)	75	75	(0.500001, 0.500001)	1.66511	8.82938e−007

Table 2

$X0$	Iter.	Call. Armijo	Opt. point	Opt. value	Riem. distance
(0.45, 0.51)	73	73	(0.499998, 0.5)	1.32776	9.75055e−007
(0.4, 0.6)	81	81	(0.499999, 0.500001)	1.32776	8.92195e−007
(0.1, 0.9)	97	97	(0.499999, 0.500001)	1.32776	9.20241e−007
(0.2, 0.3)	89	89	(0.499999, 0.499999)	1.32776	9.58094e−007
(0.7, 0.6)	84	84	(0.500001, 0.500001)	1.32776	9.98606e−007

Table 3

$X0$	Iter.	Call. Armijo	Opt. point	Opt. value	Riem. distance
(0.45, 0.51)	160	160	(0.499996, 0.500001)	1.22464	9.55101e−007
(0.4, 0.6)	178	178	(0.499997, 0.500003)	1.22464	9.4978e−007
(0.1, 0.9)	227	227	(0.499997, 0.500003)	1.22464	9.71428e−007
(0.2, 0.3)	200	200	(0.499997, 0.499997)	1.22464	9.69434e−007
(0.7, 0.6)	187	187	(0.500004, 0.500002)	1.22464	9.79192e−007

In all the numerical experiments we generate the quasiconvex function f using the composition rule $f(x) = h(g(x))$ where $g(x) = -\log(x_1(1-x_1)x_2(1-x_2))$ is a convex function on the manifold $((0,1)^n, X^{-2}(I-X)^{-2})$ and $h:\mathbb{R}\to\mathbb{R}$ is chosen as a nonconvex nondeacreasing function. We implement our code in C^{++} and all the tests were made in Pentium 866MHz with windows XP. For the implementation we use the error $\epsilon = 0.000001$ and $\alpha = 0.9$.

In Tables 1–3 above, $X0$ denotes the initial point of the algorithm, Iter. denotes the number of iterations, Call. Armijo denotes the number of Armijo test, Opt. point denotes the approximate optimal point, Opt. value denotes the approximate optimal value and finally, Riem. distance denotes the Riemannian distance between two contiguous iterations.

5.1. Experiment 1

Let $h(t) = \sqrt{t}$, then

$$f(x) = \sqrt{-\log\bigl(x_1(1-x_1)x_2(1-x_2)\bigr)}.$$

This function is quasiconvex in $((0,1)^n, X^{-2}(I-X)^{-2})$ and has a unique minimal point at $x^* = (0.5, 0.5)$ with optimal value $f^* = 2\sqrt{\log 2} = 1.665109222$.

5.2. Experiment 2

Let $h(t) = \log(1+t)$, then

$$f(x) = \log\bigl(1 - \log\bigl(x_1(1-x_1)x_2(1-x_2)\bigr)\bigr).$$

This function is quasiconvex in $((0,1)^n, X^{-2}(I-X)^{-2})$ and has a unique minimal point at $x^* = (0.5, 0.5)$ with optimal value $f^* = \log(1 + 4\log 2) = 1.32776143$.

5.3. Experiment 3

Let $h(t) = \operatorname{arctg}(t)$, then

$$f(x) = \operatorname{arctg}\bigl(-\log\bigl(x_1(1-x_1)x_2(1-x_2)\bigr)\bigr).$$

This function is quasiconvex in $((0,1)^n, X^{-2}(I-X)^{-2})$ and has a unique minimal point at $x^* = (0.5, 0.5)$ with optimal value $f^* = \operatorname{arctg}(4\log 2) = 1.224644415$.

6. Conclusions

In this work we generalize the full convergence of the steepest descent method for solving minimization problems with quasiconvex objective functions on Riemannian manifolds. In particular, this method solves constrained minimization problems with nonconvex objective functions in Euclidian spaces if the constraints constitute a complete finite dimensional Riemannian manifold with nonnegative sectional curvature and the objective function is quasiconvex in that manifold.

Observe that the sectional curvature of the manifold is used to prove that the steepest descent sequences are *quasi-Fejér convergent* and therefore bounded. Without this condition, it is possible to prove the same weak convergence result, that is, any cluster point, if it exists, is a critical point of the problem. That is useful in optimization problems on general Riemannian manifolds, for example, on Hadamard manifolds (simply connected Riemannian manifolds with nonpositive sectional curvature). Some examples of those spaces are the set of symmetric positive definite matrices $\mathcal{S}^n_{++}$, associated with the metric given by the Hessian of the barrier $-\log\det X$, and the second order cone $K := \{z = (\tau, x) \in \mathbb{R}^{1+n}\colon \tau > \|x\|_2\}$, endowed with the Hessian of $-\ln(\tau^2 - \|x\|_2^2)$.

Extension of subgradient methods to nonsmooth quasiconvex optimization problems from Euclidian to Riemannian manifolds are envisageable.

Acknowledgment

The authors are grateful to the referees helpful suggestions and useful comments for improvement of this paper, and to Alvaro Cuno Parari, PhD student at PESC, COPPE, for the insights on the numerical implementation.

References

[1] N. Barron, W. Liu, Calculus of variation l^∞, Appl. Math. Optim. 35 (1997) 237–263.
[2] A. Ben-Tal, A. Nemirovski, Lectures on Modern Convex Optimization: Analysis and Engineering Applications, MPS/SIAM Ser. Optim., vol. 2, 2001.
[3] S. Boyd, L. Vanderberghe, Convex Optimization, Cambridge Univ. Press, 2004.
[4] R.S. Burachik, L.M. Graña Drummond, A.N. Iusem, B.F. Svaiter, Full convergence of the steepest descent method with inexact line searches, Optimization 32 (1995) 137–146.
[5] J.X. da Cruz Neto, L.L. de Lima, P.R. Oliveira, Geodesic algorithms in Riemannian geometry, Balkan J. Geom. Appl. 3 (2) (1998) 89–100.
[6] J.X. da Cruz Neto, O.P. Ferreira, L.R. Lucambio Perez, A proximal regularization of the steepest descent method in Riemannian manifolds, Balkan J. Geom. Appl. 4 (2) (1999) 1–8.
[7] J.X. da Cruz Neto, O.P. Ferreira, L.R. Lucambio Perez, S.Z. Németh, Convex- and monotone-transformable mathematical programming and a proximal-like point method, J. Global Optim. 35 (1) (2006) 53–69.
[8] M.P. do Carmo, Riemannian Geometry, Birkhäuser, Boston, 1992.
[9] O.P. Ferreira, P.R. Oliveira, Subgradient algorithm on Riemannian manifolds, J. Optim. Theory Appl. 97 (1) (1998) 93–104.
[10] O.P. Ferreira, P.R. Oliveira, Proximal point algorithm on Riemannian manifolds, Optimization 51 (2) (2002) 257–270.
[11] X. Goudou, J. Munier, The gradient and heavy ball with friction dynamical systems: The quasiconvex case, Math. Program. (2007), in press; online 11 May 2007.
[12] L.M. Graña Drummond, B.F. Svaiter, A steepest descent method for vector optimization, J. Comput. Appl. Math. 175 (2005) 5–29.
[13] J. Gromicho, Quasiconvex Optimization and Location Theory, Kluwer Academic Publishers, Dordrecht, 1998.
[14] H. Wolkowicz, R. Saigal, L. Vanderberge (Eds.), Handbook of Semidefinite Programming Theory, Algorithms, and Applications, first ed., Internat. Ser. Oper. Res. Management Sci., Springer, 2005.
[15] A.N. Iusem, B.F. Svaiter, A proximal regularization of the steepest descent method, RAIRO Oper. Res. 29 (2) (1995) 123–130.
[16] K. Kiwiel, K. Murty, Convergence of the steepest descent method for minimization quasiconvex functions, J. Optim. Theory Appl. 89 (1) (1996) 221–226.
[17] S. Németh, Five kinds of monotone vector fields, Pure Math. Appl. 9 (3–4) (1999) 417–428.

[18] Y.E. Nesterov, A.N. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming, Society for Industrial and Applied Mathematics, Philadelphia, 1994.
[19] E. Papa Quiroz, P. Roberto Oliveira, A new self-concordant barrier for the hypercube, J. Optim. Theory Appl. (2007), in press; online 28 August 2007.
[20] Y.E. Nesterov, M.J. Todd, On the Riemannian geometry defined by self-concordant barrier and interior-point methods, Found. Comput. Math. 2 (2002) 333–361.
[21] T. Rapcsák, Smooth Nonlinear Optimization, Kluwer Academic Publishers, 1997.
[22] S.T. Smith, Optimization techniques on Riemannian manifolds, in: Fields Inst. Commun., vol. 3, AMS, Providence, RI, 1994, pp. 113–146.
[23] A. Takayama, Mathematical Economics, second ed., Cambridge Univ. Press, 1995.
[24] C. Udriste, Convex Function and Optimization Methods on Riemannian Manifolds, Kluwer Academic Publishers, 1994.